

SEQUENCE LISTING

SEQ ID NO:1

E. coli* expressed *P. falciparum* MSP1₄₂ (3D7)*Protein Sequence**

In Clone pET-Trx42

1 Met Ser Asp Lys Ile Ile His Leu Thr Asp
11 Asp Ser Phe Asp Thr Asp Val Leu Lys Ala
21 Asp Gly Ala Ile Leu Val Asp Phe Trp Ala
31 Glu Trp Cys Gly Pro Ccy Lys Met Ile Ala
41 Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr
51 Gln Gly Lys Leu Thr Val Ala Lys Leu Asn
61 Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
71 Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu
81 Leu Phe Lys Asn Gly Glu Val Ala Ala Thr
91 Lys Val Gly Ala Leu Ser lys Gly Gln Leu
101 Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly
111 Ser Gly Ser Gly His Met His His His His
121 His His Ser Ser Gly Leu Val Pro Arg Gly
131 Ser Gly Met Lys Glu Thr Ala Ala Ala Lys
141 Phe Glu Arg Gln His Met Asp Ser Pro Asp
151 Leu Gly Thr Asp Asp Asp Asp Lys Ala Met
161 Ala Asp Ile Gly Ser Ile Glu Gly Arg Gly
171 Thr Met Ala Ile Ser Val Thr Met Asp Asn
181 Ile Leu Ser Gly Phe Glu Asn Glu Tyr Asp
191 Val Ile Tyr Leu Lys Pro Leu Ala Gly Val
201 Tyr Arg Ser Leu Lys Lys Gln Ile Glu Lys
211 Asn Ile Phe Thr Phe Asn Leu Asn Leu Asn
221 Asp Ile Leu Asn Ser Arg Leu Lys Lys Arg

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

231 Lys Tyr Phe Leu Asp Val Leu Glu Ser Asp
241 Leu Met Gln Phe Lys His Ile Ser Ser Asn
251 Glu Tyr Ile Ile Glu Asp Ser Phe Lys Leu
261 Leu Asn Ser Glu Gln Lys Asn Thr Leu Leu
271 Lys Ser Tyr Lys Tyr Ile Lys Glu Ser Val
281 Glu Asn Asp Ile Lys Phe Ala Gln Glu Gly
291 Ile Ser Tyr Tyr Glu Lys Val Leu Ala Lys
301 Tyr Lys Asp Asp Leu Glu Ser Ile Lys Lys
311 Val Ile Lys Glu Glu Lys Glu Lys Phe Pro
321 Ser Ser Pro Pro Thr Thr Pro Pro Ser Pro
331 Ala Lys Thr Asp Glu Gln Lys Lys Glu Ser
341 Lys Phe Leu Pro Phe Leu Thr Asn Ile Glu
351 Thr Leu Tyr Asn Asn Leu Val Asn Lys Ile
361 Asp Asp Tyr Leu Ile Asn Leu Lys Ala Lys
371 Ile Asn Asp Cys Asn Val Glu Lys Asp Glu
381 Ala His Val Lys Ile Thr Lys Leu Ser Asp
391 Leu Lys Ala Ile Asp Asp Lys Ile Asp Leu
401 Phe Lys Asn Pro Tyr Asp Phe Glu Ala Ile
411 Lys Lys Leu Ile Asn Asp Asp Thr Lys Lys
421 Asp Met Leu Gly Lys Leu Leu Ser Thr Gly
431 Leu Val Gln Asn Phe Pro Asn Thr Ile Ile
441 Ser Lys Leu Ile Glu Gly Lys Phe Gln Asp
451 Met Leu Asn Ile Ser Gln His Gln Cys Val
461 Lys Lys Gln Cys Pro Glu Asn Ser Gly Cys
471 Phe Arg His Leu Asp Glu Arg Glu Glu Cys
481 Lys Cys Leu Leu Asn Tyr Lys Gln Glu Gly
491 Asp Lys Cys Val Glu Asn Pro Asn Pro Thr

10057531.01509

501 Cys Asn Glu Asn Asn Gly Gly Cys Asp Ala
 511 Asp Ala Thr Cys Thr Glu Glu Asp Ser Gly
 521 Ser Ser Arg Lys Lys Ile Thr Cys Glu Cys
 531 Thr Lys Pro Asp Ser Tyr Pro Leu Phe Asp
 541 Gly Ile Phe Cys Ser Ser *

SEQ ID NO:2

E. coli expressed *P. falciparum* MSP1₄₂ (3D7) Protein
 Sequence

In Clone pET(50)MSP1-42

1 Met His His His His His His Ser Ser Gly
 11 Leu Val Pro Arg Gly Ser Gly Met Lys Glu
 21 Thr Ala Ala Ala Lys Phe Glu Arg Gln His
 31 Met Asp Ser Pro Asp Leu Gly Thr Asp Asp
 41 Asp Asp Lys Ala Met Ala Asp Ile Gly Ser
 51 Ile Glu Gly Arg Gly Thr Met Ala Ile Ser
 61 Val Thr Met Asp Asn Ile Leu Ser Gly Phe
 71 Glu Asn Glu Tyr Asp Val Ile Tyr Leu Lys
 81 Pro Leu Ala Gly Val Tyr Arg Ser Leu Lys
 91 Lys Gln Ile Glu Lys Asn Ile Phe Thr Phe
 101 Asn Leu Asn Leu Asn Asp Ile Leu Asn Ser
 111 Arg Leu Lys Lys Arg Lys Tyr Phe Leu Asp
 121 Val Leu Glu Ser Asp Leu Met Gln Phe Lys
 131 His Ile Ser Ser Asn Glu Tyr Ile Ile Glu
 141 Asp Ser Phe Lys Leu Leu Asn Ser Glu Gln
 151 Lys Asn Thr Leu Leu Lys Ser Tyr Lys Tyr
 161 Ile Lys Glu Ser Val Glu Asn Asp Ile Lys

171 Phe Ala Gln Glu Gly Ile Ser Tyr Tyr Glu
181 Lys Val Leu Ala Lys Tyr Lys Asp Asp Leu
191 Glu Ser Ile Lys Lys Val Ile Lys Glu Glu
201 Lys Glu Lys Phe Pro Ser Ser Pro Pro Thr
211 Thr Pro Pro Ser Pro Ala Lys Thr Asp Glu
212 Gln Lys Lys Glu Ser Lys Phe Leu Pro Phe
231 Leu Thr Asn Ile Glu Thr Leu Tyr Asn Asn
241 Leu Val Asn Lys Ile Asp Asp Tyr Leu Ile
251 Asn Leu Lys Ala Lys Ile Asn Asp Cys Asn
261 Val Glu Lys Asp Glu Ala His Val Lys Ile
271 Thr Lys Leu Ser Asp Leu Lys Ala Ile Asp
281 Asp Lys Ile Asp Leu Phe Lys Asn Pro Tyr
291 Asp Phe Glu Ala Ile Lys Lys Leu Ile Asn
301 Asp Asp Thr Lys Lys Asp Met Leu Gly Lys
311 Leu Leu Ser Thr Gly Leu Val Gln Asn Phe
321 Pro Asn Thr Ile Ile Ser Lys Leu Ile Glu
331 Gly Lys Phe Gln Asp Met Leu Asn Ile Ser
341 Gln His Gln Cys Val Lys Lys Gln Cys Pro
351 Glu Asn Ser Gly Cys Phe Arg His Leu Asp
361 Glu Arg Glu Glu Cys Lys Cys Leu Leu Asn
371 Tyr Lys Gln Glu Gly Asp Lys Cys Val Glu
381 Asn Pro Asn Pro Thr Cys Asn Glu Asn Asn
391 Gly Gly Cys Asp Ala Asp Ala Thr Cys Thr
401 Glu Glu Asp Ser Gly Ser Ser Arg Lys Lys
411 Ile Thr Cys Glu Cys Thr Lys Pro Asp Ser
421 Tyr Pro Leu Phe Asp Gly Ile Phe Cys Ser
431 Ser *

SEQ ID NO:3

***E. coli* expressed *P. falciparum* MSP1₄₂ (3D7)
Protein Sequence**

In Clone pET42A

(This plasmid codes for the same peptide sequence as that
expressed in pET-IEGR-MSP1-42(AT))

1 Met Ala His His His His His His Pro Gly
11 Gly Ser Ile Glu Gly Arg Gly Thr Met Ala
21 Ile Ser Val Thr Met Asp Asn Ile Leu Ser
31 Gly Phe Glu Asn Glu Tyr Asp Val Ile Tyr
41 Leu Lys Pro Leu Ala Gly Val Tyr Arg Ser
51 Leu Lys Lys Gln Ile Glu Lys Asn Ile Phe
61 Thr Phe Asn Leu Asn Leu Asn Asp Ile Leu
71 Asn Ser Arg Leu Lys Lys Arg Lys Tyr Phe
81 Leu Asp Val Leu Glu Ser Asp Leu Met Gln
91 Phe Lys His Ile Ser Ser Asn Glu Tyr Ile
101 Ile Glu Asp Ser Phe Lys Leu Leu Asn Ser
111 Glu Gln Lys Asn Thr Leu Leu Lys Ser Tyr
121 Lys Tyr Ile Lys Glu Ser Val Glu Asn Asp
131 Ile Lys Phe Ala Gln Glu Gly Ile Ser Tyr
141 Tyr Glu Lys Val Leu Ala Lys Tyr Lys Asp
151 Asp Leu Glu Ser Ile Lys Lys Val Ile Lys
161 Glu Glu Lys Glu Lys Phe Pro Ser Ser Pro
171 Pro Thr Thr Pro Pro Ser Pro Ala Lys Thr
181 Asp Glu Gln Lys Lys Glu Ser Lys Phe Leu
191 Pro Phe Leu Thr Asn Ile Glu Thr Leu Tyr
201 Asn Asn Leu Val Asn Lys Ile Asp Asp Tyr
211 Leu Ile Asn Leu Lys Ala Lys Ile Asn Asp

221 Cys Asn Val Glu Lys Asp Glu Ala His Val
 231 Lys Ile Thr Lys Leu Ser Asp Leu Lys Ala
 241 Ile Asp Asp Lys Ile Asp Leu Phe Lys Asn
 251 Pro Tyr Asp Phe Glu Ala Ile Lys Lys Leu
 261 Ile Asn Asp Asp Thr Lys Lys Asp Met Leu
 271 Gly Lys Leu Leu Ser Thr Gly Leu Val Gln
 281 Asn Phe Pro Asn Thr Ile Ile Ser Lys Leu
 291 Ile Glu Gly Lys Phe Gln Asp Met Leu Asn
 301 Ile Ser Gln His Gln Cys Val Lys Lys Gln
 311 Cys Pro Glu Asn Ser Gly Cys Phe Arg His
 321 Leu Asp Glu Arg Glu Glu Cys Lys Cys Leu
 331 Leu Asn Tyr Lys Lys Gln Glu Gly Asp Lys Cys
 341 Val Glu Asn Pro Asn Pro Thr Cys Asn Glu
 351 Asn Asn Gly Gly Cys Asp Ala Asp Ala Thr
 361 Cys Thr Glu Glu Asp Ser Gly Ser Ser Arg
 371 Lys Lys Ile Thr Cys Glu Cys Thr Lys Pro
 381 Asp Ser Tyr Pro Leu Phe Asp Gly Ile Phe
 391 Cys Ser Ser *

SEQ ID NO:4

Nucleotide Sequence of MSP1-₄₂ (3D7) in pMBP

1

ggtaccatgg caatatctgt cacaatggat aatatacctct caggatttga
 aatgaatat gatgttatat
 atttaaaacc

81

tttagctgga gtatatagaa gcttaaaaaa acaaattgaa aaaaacattt
 ttacatttaa tttaaatttg
 aacgatatct

161

taaattcacg tcttaagaaa cgaaaatatt tcttagatgt attagaatct
 gatttaatgc aatttaaaca
 tatatacctca

aatgaataca ttattgaaga ttcatttaaa ttattgaatt cagaacaaaa
aaacacactt ttaaaaagtt
acaaatatat

aaaagaatca gtagaaaatg atatttaaatt tgcacaggaa ggtataagtt
attatgaaaa gggttttagcg
aaatataagg

atgattttaga atcaattaaa aaagttatca aagaagaaaa ggagaagttc
ccatcatcac caccaacaac
acctcogtca

```
cagcagaaaa cagacgaaca aaagaaggaa agtaagtacc ttccatTTTT
aacaacatt gagaccttat
acaataactt
```

agttaataaa attgacgatt acttaattaa cttaaaggca aagattaacg
attgtaatgt tgaaaaagat
gaagcacatg

ttaaaataac taaacttagt gattttaaag caattgatga caaaatagat
cttttttaaaa accctaccga
cttcgaagca

attaaaaaat tgataaatga tgatacgaaa aaagatatgc ttggcaaatt
acttagtaca ggattagtgc
aaattttttcc

taataacaata atatcaaaat taattgaaggy aaaattccaa gatatgttaa
acatttcaca acaccaatgc
gtaaaaaaac

aatgtccaga aaattctgga tgtttcagac atttagatga aagagaagaa
tgtaaattgtt tattaaatta
caaacaagaa

ggtgataaat gtgttgaaaa tccaaatcct acttgtaacg aaaataatgg
tggatgtgat gcagatgcca
catgtaccga

agaagattca ggtagcagca gaaagaaaat cacatgtgaa tgtactaaac
ctgattctta tccacttttc

gatggtattt

1121

tctgcagttc ctaagtcgac

SEQ ID NO:5

Primary Structure of MSP-1₄₂ in pMBP

1 Gly Ser Ile Glu Gly Arg Gly Thr Met Ala
11 Ile Ser Val Thr Met Asp Asn Ile Leu Ser
21 Gly Phe Glu Asn Glu Tyr Asp Val Ile Tyr
31 Leu Lys Pro Leu Ala Gly Val Tyr Arg Ser
41 Leu Lys Lys Gln Ile Glu Lys Asn Ile Phe
51 Thr Phe Asn Leu Asn Leu Asn Asp Ile Leu
61 Asn Ser Arg Leu Lys Lys Arg Lys Tyr Phe
71 Leu Asp Val Leu Glu Ser Asp Leu Met Gln
81 Phe Lys His Ile Ser Ser Asn Glu Tyr Ile
91 Ile Glu Asp Ser Phe Lys Leu Leu Asn Ser
101 Glu Gln Lys Asn Thr Leu Leu Lys Ser Tyr
111 Lys Tyr Ile Lys Glu Ser Val Glu Asn Asp
121 Ile Lys Phe Ala Gln Glu Gly Ile Ser Tyr
131 Tyr Glu Lys Val Leu Ala Lys Tyr Lys Asp
141 Asp Leu Glu Ser Ile Lys Lys Val Ile Lys
151 Glu Glu Lys Glu Lys Phe Pro Ser Ser Pro
161 Pro Thr Thr Pro Pro Ser Pro Ala Lys Thr
171 Asp Glu Gln Lys Lys Glu Ser Lys Phe Leu
181 Pro Phe Leu Thr Asn Ile Glu Thr Leu Tyr
191 Asn Asn Leu Val Asn Lys Ile Asp Asp Tyr
201 Leu Ile Asn Leu Lys Ala Lys Ile Asn Asp
211 Cys Asn Val Glu Lys Asp Glu Ala His Val

221 Lys Ile Thr Lys Leu Ser Asp Leu Lys Ala
 231 Ile Asp Asp Lys Ile Asp Leu Phe Lys Asn
 241 Pro Thr Asp Phe Glu Ala Ile Lys Lys Leu
 251 Ile Asn Asp Asp Thr Lys Lys Asp Met Leu
 261 Gly Lys Leu Leu Ser Thr Gly Leu Val Gln
 271 Ile Phe Pro Asn Thr Ile Ile Ser Lys Leu
 281 Ile Glu Gly Lys Phe Gln Asp Met Leu Asn
 291 Ile Ser Gln His Gln Cys Val Lys Lys Gln
 301 Cys Pro Glu Asn Ser Gly Cys Phe Arg His
 311 Leu Asp Glu Arg Glu Glu Cys Lys Cys Leu
 321 Leu Asn Tyr Lys Gln Glu Gly Asp Lys Cys
 331 Val Glu Asn Pro Asn Pro Thr Cys Asn Glu
 341 Asn Asn Gly Gly Cys Asp Ala Asp Ala Thr
 351 Cys Thr Glu Glu Asp Ser Gly Ser Ser Arg
 361 Lys Lys Ile Thr Cys Glu Cys Thr Lys Pro
 371 Asp Ser Tyr Pro Leu Phe Asp Gly Ile Phe
 381 Cys Ser Ser

SEQ ID NO:6

Nucleotide Sequence E. coli Expressed Recombinant P.
falciparum MSP-1₄₂ (3D7)

1
 atggcacacc atcatcatca tcatcccgagg ggcacccggtt ctggtaccat
 ggcaatatct gtcacaatgg
 ataatatcct

81
 ctcaggattt gaaaatgaat atgatgttat atatttaaaa ccttttagctg
 gagtatatag aagcttaaaa
 aaacaaatt

161
 gaaaaaaaca tttttacatt taatttaaatt ttgaacgata tcttaaattc
 acgtcttaag aaacgaaaat
 atttcttaga

241

tgtattagaa tctgatttaat gcaattttaa catatattcct caaatgaata
cattattgaa
gattcattta aattattgaa

321

ttcagaacaa aaaaacacac ttttaaaaag ttacaaatat ataaaagaat
cagtagaaaa tgatattaaa
tttgcacagg

401

aagggtataag ttattatgaa aagggttttag cgaaatataa ggatgattta
gaatcaatta aaaaagttat
caaagaagaa

481

aaggagaagt tcccatcatc accaccaaca acacctccgt caccagcaaa
aacagacgaa caaagaagg
aaagtaagtt

561

ccttccattt ttaacaaaca ttgagacctt atacaataac ttagttaata
aaattgacga ttacttaatt
aacttaaagg

641

caaagattaa cgattgtaat gttgaaaaag atgaagcaca tgttaaaata
actaaactta gtgattttaa
agcaattgat

721

gacaaaatag atcttttttaa aaacccttac gacttcgaag caattaaaaa
attgataaat gatgatacga
aaaaagatat

801

gcttggcaaa ttacttagta caggattagt tcaaaatttt cctaatacaa
taatatcaaa attaattgaa
ggaaaattcc

881

aagatatggt aaacatttca caacaccaat gcgtaaaaaa acaatgtcca
gaaaattctg gatgtttcag
acatttagat

961

gaaagagaag aatgtaaag tttattaaat tacaacaag aaggtgataa
atgtgttgaa aatccaaatc
ctacttgtaa

1041

cgaaaataat ggtggatgtg atgcagatgc cacatgtacc gaagaagatt
caggtagcag cagaaagaa

caggtagcag cagaaagaa

atcacatgtg

1121

aatgtactaa acctgattct tatccacttt tcgatgggtat tttctgcagt tcctaa

SEQ ID NO:7

E. coli expressed P. falciparum MSP-1₄₂ (3D7) Protein Sequence

1 Met Ala His His His His His His Pro Gly
11 Gly Ser Gly Ser Gly Thr Met Ala Ile Ser
21 Val Thr Met Asp Asn Ile Leu Ser Gly Phe
31 Glu Asn Glu Tyr Asp Val Ile Tyr Leu Lys
41 Pro Leu Ala Gly Val Tyr Arg Ser Leu Lys
51 Lys Gln Ile Glu Lys Asn Ile Phe Thr Phe
61 Asn Leu Asn Leu Asn Asp Ile Leu Asn Ser
71 Arg Leu Lys Lys Arg Lys Tyr Phe Leu Asp
81 Val Leu Glu Ser Asp Leu Met Gln Phe Lys
91 His Ile Ser Ser Asn Glu Tyr Ile Ile Glu
101 Asp Ser Phe Lys Leu Leu Asn Ser Glu Gln
111 Lys Asn Thr Leu Leu Lys Ser Tyr Lys Tyr
121 Ile Lys Glu Ser Val Glu Asn Asp Ile Lys
131 Phe Ala Gln Glu Gly Ile Ser Tyr Tyr Glu
141 Lys Val Leu Ala Lys Tyr Lys Asp Asp Leu
151 Glu Ser Ile Lys Lys Val Ile Lys Glu Glu
161 Lys Glu Lys Phe Pro Ser Ser Pro Pro Thr
171 Thr Pro Pro Ser Pro Ala Lys Thr Asp Glu
181 Gln Lys Lys Glu Ser Lys Phe Leu Pro Phe
191 Leu Thr Asn Ile Glu Thr Leu Tyr Asn Asn
201 Leu Val Asn Lys Ile Asp Asp Tyr Leu Ile
211 Asn Leu Lys Ala Lys Ile Asn Asp Cys Asn
221 Val Glu Lys Asp Glu Ala His Val Lys Ile

10054504000

231 Thr Lys Leu Ser Asp Leu Lys Ala Ile Asp
 241 Asp Lys Ile Asp Leu Phe Lys Asn Pro Tyr
 251 Asp Phe Glu Ala Ile Lys Lys Leu Ile Asn
 261 Asp Asp Thr Lys Lys Asp Met Leu Gly Lys
 271 Leu Leu Ser Thr Gly Leu Val Gln Asn Phe
 281 Pro Asn Thr Ile Ile Ser Lys Leu Ile Glu
 291 Gly Lys Phe Gln Asp Met Leu Asn Ile Ser
 301 Gln His Gln Cys Val Lys Lys Gln Cys Pro
 311 Glu Asn Ser Gly Cys Phe Arg His Leu Asp
 321 Glu Arg Glu Glu Cys Lys Cys Leu Leu Asn
 331 Tyr Lys Gln Glu Gly Asp Lys Cys Val Glu
 341 Asn Pro Asn Pro Thr Cys Asn Glu Asn Asn
 351 Gly Gly Cys Asp Ala Asp Ala Thr Cys Thr
 361 Glu Glu Asp Ser Gly Ser Ser Arg Lys Lys
 371 Ile Thr Cys Glu Cys Thr Lys Pro Asp Ser
 381 Tyr Pro Leu Phe Asp Gly Ile Phe Cys Ser
 391 Ser *

SEQ ID NO:8

tcg tcg ttt tgt cgt ttt gtc gtt

SEQ ID NO:9

PCR of genomic DNA using the forward primer

GGGGATCCATTGAGGGTCGTGGTACCATGGC AATATCTGTCACAATGG

SEQ ID NO:10

The reverse primer

GTCGACTTAGGAACTGCAGAAAATACCGG

SEQ ID NO: 11

GGGCATATGGCACACCATCATCATCATCATCCCGGGGGATCCGAC

SEQ ID NO:12

GGGCATATGGCACACCATCATCATCATCATCCCGGGGGAT CCGGTTCTGGTACCGAC